Combination of Signal Segmentation Approaches using Fuzzy Decision Making

Citation for published version:
Azami, H & Escudero, J 2015, 'Combination of Signal Segmentation Approaches using Fuzzy Decision Making', Paper presented at 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, 26/08/15 - 29/08/15 pp. 101-104.

Combination of Signal Segmentation Approaches using Fuzzy Decision Making

Hamed Azami[1,*], *Student Member*, *IEEE*, and Javier Escudero[1], *Member*, *IEEE*

***Abstract*— Segmentation is an important stage in signal analysis, and its performance plays a significant role in the efficiency of the subsequent steps, such as extraction of descriptive features and classification. There are a large number of approaches to segment signals. The performance of each of them remarkably varies when the signal changes. In this present study, two novel algorithms, which use the probability and fuzzy concepts, are proposed to combine several well-known existing signal segmentation approaches. The simulation results confirm the efficiency of the proposed approaches using the synthetic and real electroencephalogram signals.**

I. Introduction

Automatic segmentation of biomedical signals, such as electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG), is an important step in both clinical and biomedical research [1-3]. These non-stationary signals should be segmented into quasistationarity epochs so that each of them has no substantial variation in some statistical characteristics, such as mean value and variance [1]. For example, an EEG signal recorded from an epileptic patient may be divided into three segments of preictal, ictal, and postictal segments and each of which may have a different time duration [1]. There are a number of methods to segment signals [3-9].

One of the most well-known methods for signal segmentation is based on fractal dimension (FD) [3, 10]. Since FD can detect the changes in both the amplitude and frequency of the signal, it is a powerful feature extraction method for signal segmentation. There are three broadly used FD approaches, namely, Higuchi's, Katz's, and Petrosian's FD [11]. Among these methods, Katz's FD is more robust in presence of noise for EEG signals [11].

Another well-known algorithm for signal segmentation is modified Varri. This method, which is based on mean value and standard deviation of a signal, was used as a pre-processing step for classification of long-term EEG recordings [4].

A third popular signal segmentation method is based on generalized likelihood ratio (GLR) [9, 12]. In this method, two sliding windows move alongside the entire signal. In this approach, the signal content falling within each window is modeled by an auto regressive (AR) model. If the sliding windows fall within a segment, both windows will have the same statistical characteristics, and the modeling error between the two windows will be low. In contrast, if both sliding windows are not placed in the same segment, the modeling error will increase. By defining a suitable threshold level, a segment boundary point is detected when the local maximum of modeling error is above such threshold [9, 12]. To enhance this method, it was suggested to use discrete stationary WT (DSWT). This new method was called wavelet GLR (WGLR) [13].

In spite of the large number of signal segmentation approaches, each of them has its own advantages and disadvantages and its efficiency is frequently related to the statistical characteristics of signals. To address this problem, two combination algorithms based on fuzzy and probability concepts are proposed in this study. It is worth noting that the concepts of these present combination methods originated from our paper in [14] which proposed for neuronal data spike detection. Here, we will test their application in a different context, namely, signal segmentation and different kind of signals, namely, EEGs.

The synthetic signals and real EEG recordings are briefly described in Section II. Section III is devoted to the proposed methods. Section IV is concerned with simulation results and discussion. Finally, conclusions are explained in Section V.

II. Materials

A. Synthetic Signals

In this paper, 40 synthetic signals, each includes six or seven epochs with random duration between 5.5 to 8 s are used. One of the 40 signals, named test signal, is as follows:

Epoch 1: $0.5\cos(\pi t)+1.5\cos(4\pi t)+4\cos(5\pi t)$,
Epoch 2: $0.7\cos(\pi t)+2.1\cos(4\pi t)+5.6\cos(5\pi t)$,
Epoch 3: $1.5\cos(2\pi t)+4\cos(8\pi t)$,
Epoch 4: $1.5\cos(\pi t)+4\cos(4\pi t)$,
Epoch 5: $0.5\cos(\pi t)+1.5\cos(2\pi t)+0.8\cos(3\pi t)+3.5\cos(5\pi t)$,
Epoch 6: $4.5\cos(3\pi t)+2.2\cos(5\pi t)$,
Epoch 7: $0.8\cos(\pi t)+\cos(3\pi t)+3\cos(5\pi t)$.

Note that the test signal is a general and comprehensive time series because Epochs 1 and 2 are different more or less only in terms of amplitude, Epochs 3 and 4 are different nearly only in terms of frequency, and the other adjacent epochs have the different amplitude and frequency characteristics at the same time. To sum up, we have all possible states in only one signal. It is worth noting that each epoch has different random time duration.

B. Real EEG recordings

The non-invasive recording of the electrical activity of the brain over the scalp is called EEG signal. The EEG is a vital and widely used tool in clinical applications [1]. One of the most important steps of EEG analyses is signal segmentation.

[1] Hamed Azami and Javier Escudero are with the Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, King's Buildings, EH9 3JL, United Kingdom. (Phone: +44 131 650 5599, emails: hamed.azami@ed.ac.uk, javier.escudero@ed.ac.uk). * *Corresponding author*.

In this study, 40 EEG signals recorded from the scalp of ten patients are used. The length of signals and the sampling frequency are 30 s and 256 Hz, respectively. The data were prepared in the Signal Processing Research Centre at Queensland University of Technology, Queensland, Australia. The data were recorded according to the principles outlined in the Helsinki Declaration and they have been used in a number of publications about signal segmentation in the past [3, 8, 9].

III. Proposed Methods

There are a large number of signal segmentation methods. Each of them has its own benefits and drawbacks and its performance frequently varies when a signal changes. In order to overcome this problem, we propose two techniques to combine some existing methods. These proposed combination algorithms are based on the probability and fuzzy concepts to go up the accuracy of signal segmentation approaches.

A. First Proposed Method: "Hard Combination"

For each signal sample, we consider:

$$HC = \frac{\lambda_1 . SSA_1 + \lambda_2 SSA_2 + \ldots + \lambda_n SSA_n}{SSA_1 + SSA_2 + \ldots + SSA_n} \tag{1}$$

where SSA_i are signal segmentation accuracy of the i-th method. If the segment's boundary of the ith signal segmentation approach is detected $\lambda_i = 1$, else $\lambda_i = 0$. n is the number of considered signal segmentation techniques. In fact, this technique is originated from the concept of existence or absence probability of a sample as a boundary of a segment. Therefore, if this probability is more than 0.5, we assume this signal sample is a boundary of segment and vice versa. As it is clear, if SSA_m is larger than SSA_n, mth method is more reliable and trustworthy than nth method. Hence, we illustrate this effect on Equation (1).

Two parameters, including the true positive (TP) and false positive (FP) ratios, were used to assess the performance of signal segmentation algorithms. These parameters are defined as $TP = \left({N_t}/{N} \right)$ and $FP = \left({N_f}/{N} \right)$; where N_t, and N_f respectively stand for the number of correctly and falsely detected segments' boundaries and N illustrates the actual number of all segments' boundaries. It is worth noting that because the false negative ($FN = \left({N_m}/{N} \right)$; where N_m is the number of missed boundaries) used to evaluate the signal segmentation approaches is dependent on TP (TP=1-FN), we only consider TP and FP ratios here. Considering that FP is based on the inability to detect spikes, we will have $SSA = \left(\frac{TP + (1 - FP)}{2} \right)$.

For example, we investigate the influence of this technique on the test signal, described in the previous section. This is shown in Fig. 1(a). The signal's boundaries detected with modified Varri, Katz's FD and WGLR are depicted in Fig. 1(b), 1(c), and 1(d), respectively. Fig. 1 highlight some of these potential boundaries with dashed lines. In fact, every sample that at least one of the signal segmentation approaches detects as a boundary should be considered as a potential boundary. We consider the four potential boundaries marked with dashed lines, which were chosen randomly, to illustrate the behavior of the proposed method.

We combine these three approaches to evaluate the proposed method. Based on the SSAs of those methods provided in Table 1 in [3] and [3, 15] and on the fact that the signal has SNR equals to 15 dB, by replacing TP and FP values in the above equation we have SSA_1= 0.728, SSA_2 = 0.877, and SSA_3 = 0.663.

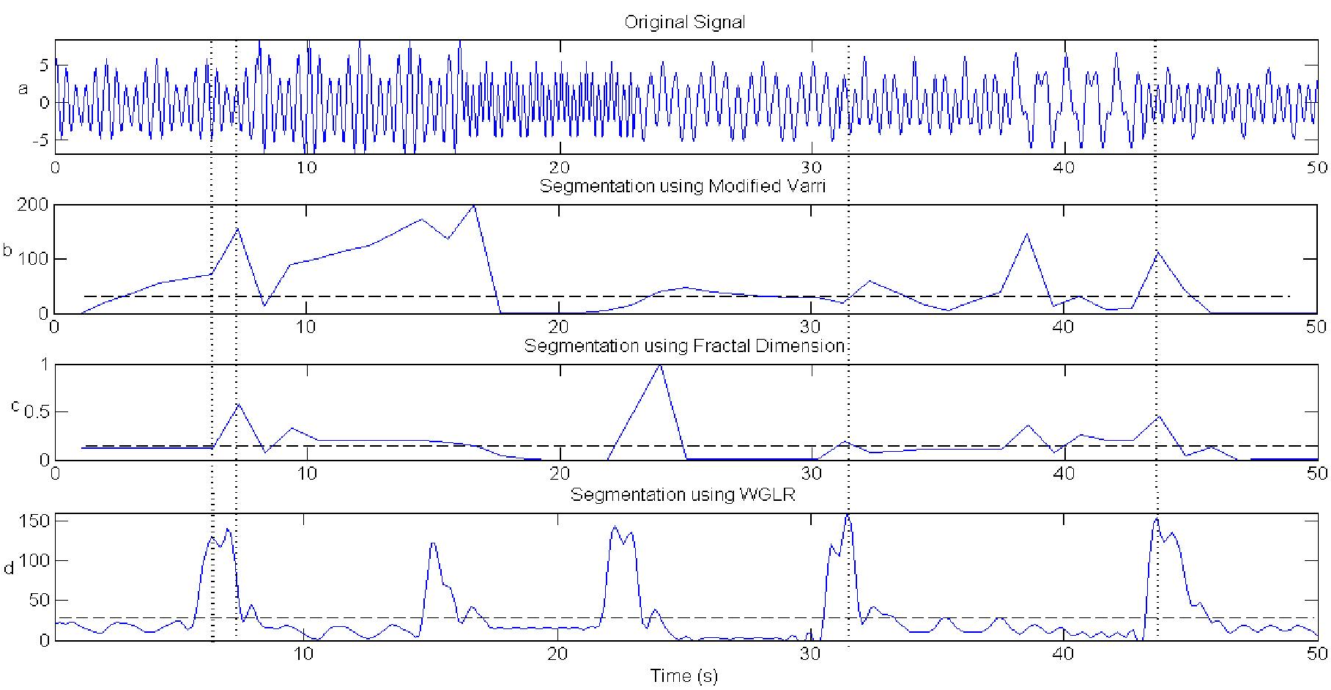

Fig. 1. The output results for combination of three existing methods; (a) original signal, (b) segmentation method based on modified Varri, (c) segmentation method based on Katz's FD, and (d) segmentation method based on WGLR.

For the first line, since $HC(1)=\dfrac{0+0+0.663}{0.728+0.877+0.663}<0.5$, the sample of the first line cannot be considered as a boundary of a segment. Considering $HC(2)$, $HC(3)$, and $HC(4)$ are more than 0.5, 2^{nd}, 3^{rd} and 4^{th} lines are segments' boundaries.

B. Second Proposed Method: "Soft Combination"

To combine some window-based existing signal segmentation methods, another method based on fuzzy and probability theory is proposed here. In this algorithm, each answer is considered as a fuzzy number between 0 and 1. Assume Fig. 2 is an output of using a window-based signal segmentation method. Beyond a doubt, for the first peak and the third peak attained by a window-based signal segmentation approach, the probabilities of being the real segments' boundaries are 1 and 0.5 respectively.

To generalize the concept, we can define two functions as follows:

$$h_{d_p}=0.5+\frac{d_p}{2d_{\max}} \tag{2}$$

$$h_{d_n}=0.5-\frac{d_n}{2d_{thr}} \tag{3}$$

where d_p and d_n are the distance between a defined threshold and a peak upper and under of the threshold respectively and h_{d_p} and h_{d_n} respectively are the fuzzy amount for a peak upper and lower than a defined threshold. It is worth noting that when $d_p=d_{max}$, then $h_{d_p}=1$ as well as if the amplitude of a peak equals with the defined threshold or $d_p=0$, then $h_{d_p}=0.5$.

Using the definition, unlike employing the conventional methods which have 0 and 1 for each peak, we can define much more precise amounts based on the fuzzy theory as follows:

$$HCF=\frac{\lambda_1.SSA_1.h_{d_1}+\lambda_2 SSA_2 h_{d_2}+\ldots+\lambda_n SSA_n.h_{d_n}}{SSA_1+SSA_2+\ldots+SSA_n} \tag{4}$$

where h_{d_i} is $h_{d_{pi}}$ or $h_{d_{ni}}$ when the peak is higher or lower than the defined threshold, respectively.

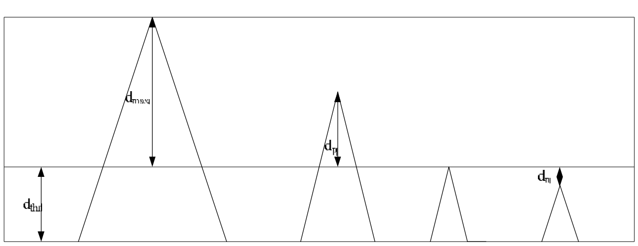

Fig. 2. The output when a window-based signal segmentation method is employed.

It worth noting that this kind of definition is useful for combination of some signal segmentation techniques. As an example, for the first line of Fig. 1, we have:

$$HCF(1)=\frac{0+0+1\times 0.663(0.5+\frac{130}{2\times 155})}{0.728+0.877+0.663}<0.5$$

Therefore, the sample of the first line cannot be considered as a boundary of a segment in this approach. Likewise the results of the first algorithm for combination of signal segmentation methods, $HCF(2)$, $HCF(3)$, and $HCF(4)$ are higher than 0.5. Hence, the 2^{nd}, 3^{rd}, and 4^{th} lines can be considered as segments' boundaries.

IV. SIMULATION

A. Synthetic Signals

In this paper, 40 synthetic signals are used to evaluate the capability of the proposed methods in comparison with three well-known signal segmentation approaches. This is shown in Table I. As can be seen in Table I, among existing methods, modified Varri and WGLR are the best and worst ones in terms of both TP and FN. Regarding FP, the FD-based method has better signal segmentation accuracy. In terms of all evaluating parameters, the first and second proposed methods are remarkably better than all the existing methods although the second combination technique is slightly better than the first one in terms of FP.

B. Real EEG Data

Table II illustrates that the proposed methods, compared with the three other well-known existent methods, namely, FD-based, WGLR, and modified Varri, have better TP, FN, and FP ratios for segmenting the 40 real EEG recordings. Considering the two proposed methods, the TP and FN ratios are slightly better for the first proposed method, even though FP ratios for the second approach slightly better. Among the three existing methods, FD-based one has higher accuracy in all evaluating metrics. Albeit the modified Varri has better TP and FN ratios than WGLR, its FP ratio is worse than that of WGLR. Furthermore, the WGLR method does not have acceptable TP and FN ratios. Hence, this method is not reliable.

TABLE I. SIGNAL SEGMENTATION RESULTS OF THE AFOREMENTIONED IMPROVED AND EXISTING METHODS USING 40 SYNTHETIC SIGNALS.

Method	TP	FN	FP
Signal segmentation based on FD [3, 10]	93.2%	6.8%	2.9%
Signal segmentation based on modified Varri [4, 15]	100%	0%	54.4%
Signal segmentation based on WGLR [9, 13]	45.6%	54.4%	13%
The first proposed combination approach	100%	0%	2.1%
The second proposed combination approach	100%	0%	2%

TABLE II. SIGNAL SEGMENTATION RESULTS OF THE AFOREMENTIONED IMPROVED AND EXISTING METHODS USING 40 EEG SIGNALS.

Method	**TP**	**FN**	**FP**
Signal segmentation based on FD [3, 10]	89.8%	10.2%	8.1%
Signal segmentation based on modified Varri [4, 15]	87.1%	12.9%	60.2%
Signal segmentation based on WGLR [9, 13]	67.8%	32.2%	38%
The first proposed combination approach	94.3%	5.7%	7.5%
The second proposed combination approach	93.9%	6.1%	6.9%

V. CONCLUSION

The present study shows the importance of combination of results obtained by different signal segmentation approaches. We have proposed two techniques based on the probability and fuzzy concepts and tested these methods using a set of synthetic signals and EEG data. Both proposed methods have had relatively similar results and have been significantly better than the existing signal segmentation approaches. We have achieved 100% and about 94% TP ratios on sets of 40 synthetic signals and EEGs, respectively, using proposed combination approaches. In terms of TP and FN, the first proposed method has achieved better results than the second one using EEG signals, although for synthetic data both have had the same performance. In terms of FP, the second proposed technique has been better than the first one using both synthetic data and EEG recordings.

ACKNOWLEDGMENT

The authors are grateful to Prof. Boualem Boashash from the College of Engineering, Qatar University, Qatar, and Prof. Hamid Hassanpour from the Department of Information Technology and Computer Engineering, Shahrood University of Technology, Iran for providing and giving the real EEG signals.

REFERENCES

[1] S. Sanei and J. Chambers, *EEG signal processing*: John Wiley & Sons, 2008.

[2] P. Micó, M. Mora, D. Cuesta-Frau, and M. Aboy, "Automatic segmentation of long-term ECG signals corrupted with broadband noise based on sample entropy," *Computer Methods and Programs in Biomedicine,* vol. 98, pp. 118-129, 5// 2010.

[3] H. Azami, S. Sanei, K. Mohammadi, and H. Hassanpour, "A hybrid evolutionary approach to segmentation of non-stationary signals," *Digital Signal Processing,* vol. 23, pp. 1103-1114, 2013.

[4] K. Kosar, L. Lhotska, and V. Krajca, "Classification of long-term EEG recordings," in *Biological and Medical Data Analysis*, ed: Springer, 2004, pp. 322-332.

[5] T. K. Truong, C.-C. Lin, and S.-H. Chen, "Segmentation of specific speech signals from multi-dialog environment using SVM and wavelet," *Pattern Recognition Letters,* vol. 28, pp. 1307-1313, 8/1/ 2007.

[6] R. Agarwal and J. Gotman, "Adaptive segmentation of electroencephalographic data using a nonlinear energy operator," in *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on*, 1999, pp. 199-202 vol.4.

[7] P. Mazurkiewicz, "Automatic Segmentation of EMG Signals Based on Wavelet Representation," in *Computer Recognition Systems 2*. vol. 45, M. Kurzynski, E. Puchala, M. Wozniak, and A. Zolnierek, Eds., ed: Springer Berlin Heidelberg, 2007, pp. 589-595.

[8] H. Azami, H. Hassanpour, J. Escudero, and S. Sanei, "An intelligent approach for variable size segmentation of non-stationary signals," *Journal of Advanced Research,* 2014.

[9] H. Azami, H. Hassanpour, and S. Anisheh, "An Improved Automatic EEG Signal Segmentation Method based on Generalized Likelihood Ratio," *International Journal of Engineering-Transactions A: Basics,* vol. 27, p. 1015, 2014.

[10] M. E. Kirlangic, D. Perez, S. Kudryavtseva, G. Griessbach, G. Henning, and G. Ivanova, "Fractal dimension as a feature for adaptive electroencephalogram segmentation in epilepsy," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 2001, pp. 1573-1576 vol.2.

[11] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on,* vol. 48, pp. 177-183, 2001.

[12] U. Appel and A. V. Brandt, "Adaptive sequential segmentation of piecewise stationary time series," *Information sciences,* vol. 29, pp. 27-56, 1983.

[13] L. Jun, L. Xing, and L. Tong, "Web-Based Application for Traffic Anomaly Detection Algorithm," in *Internet and Web Applications and Services, 2007. ICIW '07. Second International Conference on*, 2007, pp. 44-44.

[14] H. Azami, J. Escudero, A. Darzi, and S. Sanei, "Extracellular spike detection from multiple electrode array using novel intelligent filter and ensemble fuzzy decision making," *Journal of Neuroscience Methods,* vol. 239, pp. 129-138, 1/15/ 2015.

[15] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An improved signal segmentation using moving average and Savitzky-Golay filter," *Journal of Signal and Information Processing,* vol. 3, pp. 39-44, 2012.